DOCUMENT RESUME

ED 083 810 EM 011 531

AUTHOR	Dostert, Bozena Henisz
TITLE	REL--An Information System for a Dynamic Environment.
INSTITUTION	California Inst. of Tech., Pasadena.
SPONS AGENCY	National Institutes of Health (DHEW), Bethesda, Md.
REPORT NO	CIT-REL-R-3
PUB DATE	Dec 71
NOTE	50p.; See also EM 011 530 and EM 011 532 through EM 011 536
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Computer Programs; Computers; Computer Science; Data; Data Bases; Data Processing; *Interaction; *Man Machine Systems; Programing; *Programing Languages; Technical Reports; Time Sharing
IDENTIFIERS	Interrelated Data; Natural Language; *Rapidly Extensible Language; REL; Time Oriented Data

ABSTRACT

Three outstanding features of the Rapidly Extensible Language (REL) System are discussed. REL itself is an integrated software system designed to facilitate conversational interaction with the computer, especially by those working with dynamic, highly interrelated data. The system's main parts are: 1) the operating system, which manages the simultaneous use of the system from a number of terminals and handles all input/output from peripheral storage; 2) the language processor, which analyzes the incoming query or data and schedules and executes the appropriate calculations and processing of the data base; and 3) the REL languages and user language/data base packages. The characteristics of the system treated in detail in this paper include: 1) its ability to handle interrelated and time-oriented data; 2) its provision for communication with data in natural language trailored to the user's needs, with emphasis upon ordinary English; and 3) its extensional facility, which allows for the modification of data through definitions of new terms and relationships as part of the user's ongoing work with the system. (Author/LR)

ED 083810

a project report on

REL

Co-principal investigators:
Bozena Henisz Dostert
Frederick B. Thompson

California Institute of Technology
Pasadena, California, 91109

EM 011 537

FILMED FROM BEST AVAILABLE COPY

ED 083810

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

REL -- An Information System for a Dynamic Environment

Bozena Henisz Dostert

REL Report No. 3

California Institute of Technology

Pasadena, California

December, 1971

The research reported here was

supported in part by the National

Institutes of Health, grant GM 15537

REL -- An Information System for a Dynamic Environment

I. Introduction

REL stands for Rapidly Extensible Language System which is being developed at the California Institute of Technology. REL is an integrated software system designed to facilitate conversational interaction with the computer especially on the part of those working with dynamic, highly interrelated data, in situations where the data is not only to be accessed but also to be manipulated in various ways by the user to suit his specific needs. In such situations, the user must be able to work with his data in a natural manner, through a language that is natural to him and best suited to his task, a language that allows him to analyze the data in a most facile and meaningful way. An essential ingredient of such a language is its extensibility, the ability to define and redefine terms so as to find the essential interrelationships in the data.

Most current data management and analysis systems are built around the notion of the total management information system. In such systems, data is collected from all over a large organization, stored in a large and all-encompassing data base and made available to higher levels of management through statistical analysis routines or report generators. To attain the necessary levels of efficiency, the operation of such a system must be centralized; to remain sufficiently stable to be useful to the management, the content and format of the material must be closely regulated and under the control of the information system operation.

But there are other kinds and uses of data in organizations, whether they be a research center, an industrial set up, or a military headquarters. Each research team or staff office has files of information of their own which they use constantly as an ongoing part of their work. These may be records and results of a current series of experiments, or the data and models the team is working with in putting together a special study, or working files of raw material relating to ongoing research, or records on alternative budgets and planning charts used in the preparation of a new program proposal. In all such cases, the research team or staff office is directly involved in gathering and maintaining this material, in making day to day decisions on its contents, formats and file organizations. Such material is not appropriate for the master file of the larger or outer organization and it is far too dynamic in all its dimensions for standardization. Further, those who develop and use such materials would not think of giving up control over them, for they are in a real sense the stuff and substance of their ongoing work. These are the dynamic, working files that constitute the essence of research and staff operations.

In order to build a system which is responsive to the needs of users in such dynamic environments, we have been especially attentive to two characteristics of the work of individual users or groups who analyze different aspects of a body of data much of which may be common to several individuals or groups. First, they need to deal with their data in an individualized manner, to dissect it in new ways, to test even far-fetched hypotheses, to build up their terminology in order to deal with the data most efficiently. Second, they need to communicate with each other's data, consult, and benefit from each other's analyses.

What, then, should be the characteristics of an information system

which aims to support the working files, and habits, of many staff offices and research teams in dynamic environments?

First, such a system has to have the capacity to handle highly interrelated and time-oriented data. It must allow individual queries and analyses along unanticipated avenues, and allow for the tracing of complex interrelationships. The very essence of research and staff studies lies in the search for new interrelationships, following of clues, and even guesses, tracing of implications and clarifications of emerging patterns. Thus, a system designed for supporting such operations must facilitate innovative, unprogrammed exploration of the data.

Second, in such an information system, communication between the user and his data must be in a language natural to him and tailored to his needs. On the one hand, this requirement calls for man/machine languages built on the syntax of natural language. On the other hand, the vocabulary and idioms of such languages should be those of the working teams or individuals, they should reflect the idiosyncratic dialect that is built around the concepts and interrelationships relevant to their work.

Third, such a system must be able to accept new terms and new data, as well as new definitions of functions and relationships in the process of the ongoing use of the system; and it must incorporate these language and data base extensions for immediate use and further extensions. It is this changing, dynamic character that is essential to the work of the staff, since the modifications in the data base and the concomittant modifications of the language are a reflection of the staff's maintenance of the relevance of the data, and thus a reflection of how well they are doing their job.

Fourth, such an information system must provide enormous flexibility, which allows for a variety of user language/data packages, individually tailored to specific needs, and provide a facility for the intercommunication

of such specialized packages. It must allow the addition of new language/data packages and new algorithms, and the employment of a wide variety of data structures.

Fifth, but that goes without saying, such a system must have good response times.

The REL system is designed to fulfill all of the above requirements. It has already stood the test of users with needs such as were discussed above. This experimental REL system was in operation in the spring and summer of 1970. We are now developing a fully operational prototype. This prototype will be a fully interactive, multiprogrammed system by which a number of researchers can communicate directly with their data and models in a conversational way, time-sharing the computer facilities. This prototype system should be in operation a year from now on an IBM 370/135 in a test environment. We plan for a debugged, evaluated and documented system by the fall of 1973.

In this paper, only some of the outstanding features of the system are discussed and illustrated, namely:

(1) its ability to handle interrelated and time-oriented data;

(2) provision for communication with data in natural language, tailored to user's needs, with emphasis upon ordinary English;

(3) the extensional facility, which allows for the modification of data through definitions of new terms and relationships as part of the user's ongoing work with the system.

The third point receives special emphasis, since its discussion and illustration also serves to bring out the other features.

First, however, the general architecture of the system needs a brief presentation. It is more fully discussed in [1,2,7].

Design of the REL System

The REL system has three main parts. As Figure 1 shows, the three main parts are:

(1) the operating system, which manages the simultaneous use of the system from a number of terminals and handles all input/output from peripheral storage,

(2) the language processor, which analyzes the incoming query or data, and schedules and executes the appropriate calculations and processing of the data base,

(3) REL languages and user language/data base packages.

One of the most distinguishing features in the architecture of REL as compared with other relational data systems is that it has a single language processor for all languages, and that this language processor is tightly coupled with the operating system. In most computing language systems, the system can accept and process statements of a given language by using a separate compiler that is specific to the particular language. REL, however, has a single language processor that can handle a wide variety of high level languages. In essence, this language processor is a straight forward syntax directed interpreter. It includes built-in facilities for handling variables and recursion, and provides for extensions by users of the languages.

This feature of REL architecture has several major advantages. First, it is much easier to implement a new language or extend an existing one. Languages can be conveniently tailored to particular applications and specialized processes can be added to one's language as the need arises.

Second, since this single language processor is closely tied to the underlying operating system, it allows efficient scheduling, allocation

and access of peripheral storage which could not be achieved otherwise.

Third, in applications where a number of offices or groups each have their own "system", that is a language/data base package, the specific architecture of REL facilitates inter-communication between such subsystems.

The technical problems of implementing a natural language question-answering system are quite different from those encountered in programming language compilers. From the system programmer's point of view the relevant characteristic of most REL applications is that they deal with large data bases that must be kept in disk memory. The prime problem is efficient access to that memory. One solution is to restrict the nature of the questions that can be asked and optimize disk access methods around these restricted queries. This solution is not acceptable in the majority of applications to be found in dynamic situations. The REL solution, and a principal element of the system, are the paging algorithms for the dynamic optimization of access to the disk memory in terms of the data requirements of each query.

The language writer controls both the allocation of data to individual pages and the page segmentation of the interpretive routines, and he can do this without becoming involved in the details of the language processor or the paging mechanism. As a result, there is a rational relationship between lexicon and syntax, on the one hand, and the allocation and retrieval of pages from disk storage, on the other. Scattering of data and routines haphazardly over the peripheral storage, a source of major inefficiencies in other systems, is avoided.

One other design feature must be mentioned in connection with processing of data with complex interrelationships, which experience with the system in the summer of 1970 pointed to. An investigator who has a complex data base is soon led to ask questions that call for an extensive

amount of computation and data manipulation. Such an investigator is usually well aware that he must wait a considerable time for his answer, and since he is aware of the amount of computation he has asked for, is prepared for the delay in the response. This use pattern is likely to be quite typical, in a system where the user's language can be so easily extended, thus providing the means of succinctly expressing complex questions.

To facilitate this pattern of usage, REL will have the capability to cast off a query into the "background" as a low priority job in the system, and free the terminal for continued conversational use in the interim period. Since the REL operating system is a multiprogrammed system in which several jobs are resident in core memory at the same time, each occupying one of the available "slots", the plan is to make one dynamically allocated slot available for background jobs. Thus, one could say that the system would be both an interactive system and a conversationally driven batch system.

Such details of implementation have to be mentioned, in view of our over-all objectives.

To turn now to languages within the system, REL languages are of two types, which we call "base" languages and "user" languages. A highly specialized "user" language can be developed for a particular user, incorporating the syntax and basic algorithms natural to his problem area. More commonly, however, a user will make use of a general language already available, tailoring it to his own needs by introducing his own vocabulary and definitions.

Two such "base" languages have been implemented and applied by users. One is REL English and the other the REL Animated Film Language.

REL English,further discussed and illustrated in Section III, is a

sizable subset of natural English. In the base version, the vocabulary is limited to the "little" words such as "all", "and", "what", "before". Together with the grammar rules for natural English, this constitutes a base on which a user can build his own special language, and then extend it and modify it according to his needs. As he makes use of the inherent definitional capability of the system, his language and his data base become tightly interwoven, constituting his own language/data base package. Several actual examples from user experience with REL English and the construction of a specialized language/data base package are given in Section III.

Now, then, what exactly do we mean when we refer to the "rapid extensibility" of REL languages? Our notion of extensibility derives from our understanding of how a researcher or any person dealing with dynamic data goes about his work. As his understanding of his material grows, he develops new concepts, finds new patterns in his data, interrelates his data in new ways. This evolving conceptualization is mirrored in his use of language. He defines new patterns and relationships in terms of old, and adds terms as he needs them. As he moves forward, he makes use of those newly defined terms and concepts. In dealing with his data, he needs to be able to communicate with the computer in these new terms rather than always having to express himself in some rudimentary language. Only in this way can he use the computer as a facile tool of his analysis.

The REL definitional capability can best be illustrated through the experience of two users who worked with those two entirely different languages: REL.AFL, the animated film language, and REL English.

Mr. John Whitney, the computer artist of international renown, used the REL.AFL language. This language is a highly specialized language for conversational interaction with the graphic display terminal (IBM 2250)

for composition and subsequent animation of motion picture films. Mr. Whitney used this language to make his film called MATRIX, which he presented at the International IFIP 71 Congress in Ljubljana, Yugoslavia, last August. [3]

In a typical working session, Whitney could define several visual forms, say a cube and a pleasing space curve. The definition of these forms might be either in terms of an array of simpler forms, e.g line segments arranged to define the planar projection of the cube, or might involve mathematical expressions, for instance in defining the space curve. Once defined, these forms could then be manipulated by the artist as conceptual units and be composed into higher level forms and sequences. For example, a series of cubes might move rhythmically along the space curve in such a way as to move into, and out, of symmetric interrelationships. The artist would then proceed by executing and modifying his developing composition on the display scope, working with the visual images to bring the ultimate composition into artistic balance. If it had been necessary for him to state these higher level compositions in terms of the basic shapes of two-dimensional lines and mathematical equations, rather than in terms of cubes and the space curve, the artist would have been strained beyond his ability to conceptualize.

As far as REL English is concerned, this passage to new, high level conceptual forms can be seen in the protocol of Dr. Thayer Scudder who made extensive use of the experimental REL system. Dr. Scudder, a Caltech anthropologist, and Dr. Elizabeth Colson, of the University of California at Berkeley, used the REL system to analyze their data concerning the Gwembe Tonga, a people living in Zambia. Their data base was of the order of 10^5 items. The following illustration is from one of Dr. Scudder's sessions with the computer. First he defined the term "sex ratio". Later

on, he was interested in considering only the older women of the Mazulu village, whom he defined as "Mazulu crones". He could then ask:

What is the sex ratio of the children of Mazulu crones?

instead of

What is the number of male children of Mazulu dames who were born before 1920, times 100,divided by the number of female children of Mazulu dames who were born before 1920?

On the surface, this seems a minor advantage. However, in the process of on-going investigation, the recognition, testing and establishment of new conceptual forms is expected to take this step-by-step path. These steps, as they build up, evolve into new and more revealing conceptual patterns. How a user extends his language through definitions during his on-going conversation with the computer is well illustrated through Scudder's protocol. Definitions can, of course, be deleted and changed, as well as added. The concepts defined, as well as the questions that can be asked, may involve higher level abstractions and complex interrelationships, not just simple identifiers of individual entities or subsets of the data, as might seem initially.

As a user builds up a hierarchy of definitions, computing efficiency is likely to be degraded when the higher forms are used, especially when they entail complex calculations on the data. The investigator should then have recourse to a programming staff who can replace the hierarchy of definitions leading to a term by an efficient algorithm expressing internally the complex meaning of the term.

Thus, REL provides for two kinds of language extension. First, it is easy for the investigator himself to define new terms and extend and modify his language, i.e. his lexicon and his data, while working with the data. Second, it is easy, at the programming staff level, to initiate

and extend languages tailored to the needs of the users. It is precisely these two capabilities that constitute the extensibility provided by REL.

Finally, the data itself may need frequent extensions. There are two sides to this issue: adding small amounts of data, which the investigator can add just as easily as he adds definitions. Such additions are immediately incorporated into the data. The other side is adding large bodies of data, particularly when that data is on punch cards in typical field formatted form. Let us consider, for example, a data deck whose card format is:

NAME	POPULATION	LAND AREA
France	45540	213

Using a language based on REL English, an investigator could enter from a terminal the following definition:

```
def: "France" "45540" "213":
The population of "France" is "45540" and the land area of
"France" is "213".
```

The quotes indicate that any other similar term may be used in place of the ones shown, e.g. names of other countries and other numerical data. It is easily noticed that this simple definition decodes the card format into a statement whose processing will build the facts indicated into the data base. Having submitted the card deck to the machine operator, one types: "Alternate input: cards." and the system then processes the data cards, whose translation is understood by the language processor in terms of the above definition.

III. REL English

REL English is currently the most prominent language within the REL system, one that has already been tested extensively in user applications

and one that can find a variety of applications as a natural means of communication with the computer.

Just as in ordinary English we use different modes of expression, different styles to suit specific situations, in REL English not all constructions of ordinary English are available. For instance, colloquial, casual on the one hand, and extremely elaborate constructions on the other, are not part of REL English. However, we are continually bringing it closer to normal English by incorporating new structural features. Currently, REL English grammar consists of over 350 rules which allow a variety of constructions to be handled. The grammatical structure of REL English is discussed in [4,5,6], and here the presentation is limited to illustration of the constructions that can be handled and samples of actual conversations with the data.

As for the range of constructions handled, REL English uses

(a) complex verb structures, including references to time e.g.

Had John been given the message before his Boston friend arrived?

Did John arrive in New York after July 1, 1970?;

(b) relative clauses, e.g.

Did some boy see the girl who left London?

Did John give Mary books which he bought from Tom?;

(c) complex noun phrases, e.g.

Mary is the daughter of John's wife's brother.

John sent a letter to his wife's mother.

(d) quantifiers, which select data and group it, e.g.

Which ships left Boston after May 1971?

How many reports were sent by John last year?

(e) conjunctions, which join nouns and sentences, e.g.

Did John live in New York or Boston?

Mary attended Harvard and her brother enrolled in Yale.

Combinations of such constructions, with some others not illustrated here, make it possible to use REL English with ease and a feeling of conversation in natural English. Their use can best be observed in the following samples of conversations. The section marked A contains queries put to a small experimental ficticious data base containing information on a few families, their dates of marriages, children's birthdates, and incomes. These conversations were conducted in July, 1970. Section B are Dr. Scudder's protocols.

A. What is the location of Waverley's author?

Edinburgh

def:arrive:=verb(location, 1)

def:depart:=verb(location, 2)

def:live:=verb(location, 0)

Where did Sir Walter Scott live?

Edinburgh

Notice here the definition of the verbs"arrive", "depart", and "live" in terms of "location". The digits 1, 2, 0 refer to the beginning, ending and duration of the location. A user, given a manual of some simple rules, can thus add even verbs to his language. But the expectation is that users will choose a base language which already contains an extensive number of verbs.

When did Sally Smith live in New York?

from April 1935 to August 1955

from September 1957 to June 1970

Where was Sally Smith in 1956?

Washington

Notice here that since Sally lived in New York at different times, the specific periods are given.

Who is Jill's sister-in-law?

Eh?

def:sister-in-law:sister of spouse

Here we see the introduction of a new term which did not exist in the language by defining it through already existing terms. The old terms may themselves be the result of earlier definitions, e.g. "sister" would result from

def:sister:female sibling

and "sibling" itself is the result of a definition. Again, the user may introduce his own layers of definitions, or choose a base language already containing those he might find useful.

The next sentence shows how, once a new term has been introduced, it can be used immediately in further exploration of the data

When was Bill Barton's wife Jill's sister-in-law?

from September 1958 to June 1970

Next, we observe the use of the quantifiers "each" and "who" which result in listings of selections from the data, and of the use of time information. The last answer shows that Sandra Smith was a Smith only until September 1956, when she became a Barton.

When did each male Smith live in each city?

Sam Smith	New York	from 1930 to September 1955
	Washington	from October 1955 to June 1970
Stan Smith	Los Angeles	from June 1961 to June 1970
	New York	from December 1933 to April 1961
Steve Smith	Los Angeles	from July 1962 to June 1970

When was the number of Smiths equal to 5?

from August 1938 to September 1956

from September 1958 to April 1960

Who were Smiths in 1956 and 1959?

Ann Andersen Smith

Jill Jones Smith

1959

Sally Smith

Sam Smith

Sandra Smith Barton

from January to September 1956

The use of the conjunction "and" can be seen above in the phrase "in 1956 and 1959". And also in the following sentences

Did Stan marry Jill, and live in Los Angeles?

Yes

Did Sam Smith live in some eastern city and Los Angeles?

eastern city	Yes
Los Angeles	No

The following question again illustrates the manipulation of time-related data through the use of a subordinate clause referring to time, viz. "since Steve Smith was born".

What were the incomes of the parents of Steve Smith

since Steve Smith was born?

Jill Jones Smith	4000 from July to December 1962
Stan Smith	8000 from July to December 1962
	9000 in 1963
	10000 from 1964 to 1966
	11000 from 1967 to 1969

It will be noticed that REL English in July 1970 did not yet handle pronouns. Therefore the repetition of "Steve Smith" instead of "he".

If the above question were phrased "What was the income...", the sum

of incomes would appear for the year 1962, as the next question illustrates.

What was the income of Joneses in 1956?

19000

What was the average income of Joneses in 1956?

4750

What was the maximum income of Joneses in 1956?

10000

Higer level calculations can be performed easily since REL English contains a strong mathematical component.

Definitional capabilities are further demonstrated in the following sequence

What is the definition of aunt?

1. aunt:sister of parent

What is the definition of sister?

1. sister:female sibling

What is the definition of sibling?

1. sibling:child of parent, but not identity

What is the definition of child?

1. child:converse of parent

What is the definition of parent?

1. parent:=relation

Who is Steve Smith's grandfather?

Eh?

def:grandfather:father of parent

Who is Steve Smith's grandfather?

John Jones

Sam Smith

Notice that the nature of a definition can always be recalled simply by asking for its content.

The final sequence relating to this data base shows how new data can be incorporated.

Who are Jill's children?

Steve Smith

Sue Smith

def:Ted Smith:=name

Ted Smith is a male!

Ted Smith was born in June 1969!

The parents of Ted Smith are Jill and Stan!

Who are Jill's son's?

Steve Smith

Ted Smith

B.

The protocols which follow were made by Dr. Thayer Scudder, the Caltech anthropologist. Mazulu is the name of one of the villages studied.

What is the number of total 1957 Mazulu sample?

138

What is the number of 1957 Mazulu sample who are male?

64

What is (64/(138-64)) * 100?

86.4864865

def:sex ratio of "Mazulu samples":(number of "Mazulu samples" who are male) * 100/(number of "Mazulu samples" who are female)

What is the sex ratio of total 1970 Mazulu samples?

85.6060606

The above sequence, and the following, again illustrate the process of definitional build-up to serve novel explorations of the data.

What is the number of Mazulu dames57?

31

def:Mazulu crone:=name

All Mazulu dames57 who were born before 1920 are Mazulu crones!

What is the number of Mazulu crones?

9

What is the sex ratio of children of Mazulu crones?

74.2857143

To quote Scudder's own commentary:

> "A major attribute of REL encourages the investigator to create definitions as he proceeds. On page 1 sex ratio is defined, while at the bottom of the page a category "Mazulu crones" is defined in such a way as to include only women who have completed their reproductive span. Since any number of such categories can be defined, the investigator can compare, for example, age grades in regard to whatever traits he wishes. Because of the ease with which definitional categories can be created, the user, for the first time, has the opportunity to "play around" with his data--to look for relationships which have yet to be explicitly conceptualized. This is a tremendous advantage in social anthropology with its incredibly rich informational base. REL facilitates not only the testing of old theories but also, through rapid scanning of much data, the generation and testing of new theory." [7]

Dr. Scudder's commentaries below the next two passages show the reaction of a user as he analyses his data with the help of building definitions.

def:polygyny rate "in 1957" of "people":(number of marriages "in 1957" of "people")/(number of "people" who were married "in 1957")

What is the polygyny rate in 1970 of Mazulu machismos57?

1.84

"Once polygyny rate is defined, it is, of course, easy to compare polygyny rates of the same aggregate of men at different points in time as well as those of different age categories at any given date." [7]

What were the villages of major upbringing of 1957 wives of Mazulu machismos57?

chisamu

makololo

masili

mazulu

mukwapata

munjola

nashongo

pembazana

siakaloba

siamwalu

simuchembu

sinafwala

unknown village

Mazulu was the village of major upbringing of which 1957 wives of Mazulu machismos57?

Chitugo Kombo

Impunga Muntanga

Kajamba Simakulita

Kapule Mukuli

Malimbu Mulongo

Munsaka Kalinda

Siankumba Changu

Sophia Miyoba

"A logical sequel to the questions dealing with village of wife's major upbringing would be to define the largest indigenous political unit among the Tonga (the cisi or neighborhood) in terms of the specific villages which make it up. For example, def:miyaka cisi:=mazulu pembazana chisamu and siamyalu villages. One could then ask what proportion of 1957 (etc.) wives of Mazulu machismos57 (etc.) were born in miyaka cisi. In this case the answer would be over 80%". [7]

def:shade holder:=name

All Mazulu machismos57 who had shades are shade holders!

What is the number of shade holders?

7

What is the definition of igobelo?

1. igobelo:converse of homestead

What is the number of 1957 igobelos of Mazulu Josam?

7

What is the number of 1957 igobelos of each shade holder?

Mazambani Sitente	1
Siacaligonya	37
Sialozi Siabenzu	8
Tembo Siankede	9

What was the polygyny rate in 1957 of shade holders?

2.142857143

One of Sucdder's comments was that, "a great strength of REL is that the investigator can afford, in playing with the data, to search out a multiplicity of relationships whereas in using other techniques he might settle on a single suspected relationship and after lengthy statistical analysis be tempted to read too much into correlations found." [7]

Comments such as Dr. Scudder's make those of us in the design and implementation of the REL system and REL English feel that we have already had very valuable experience with the system and we are confident of the REL promise.

With the tremendous developments in computing which we have witnessed in the past two decades, it is now time that computers should be "humanized" and that many men, and women, be liberated from the distance between men and machines. Computers should be easily manipulatable tools in the hands of those to whose work they could contribute immensely--members of dynamic, complex environments. REL is a computer system for these types of users.

REL Architecture

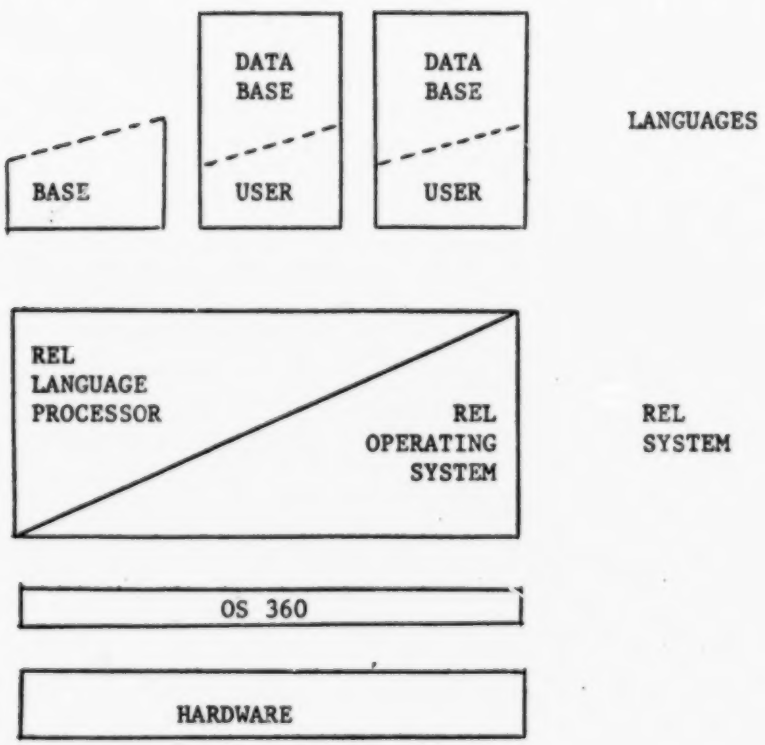

FIGURE 1.

REFERENCES

1. Thompson, F.B., Lockeman, P.C., Dostert, B.H., and Deverill, R.S., "REL: A Rapidly Extensible Language System," Proc. 24th National ACM Conference, August, 1969.

2. Lockemann, P.C., and Thompson, F.B., "A Rapidly Extensible Language System: The REL Language Processor", 1969 International Conference on Computational Linguistics, Stockholm, 1969.

3. Whitney, John H. "A Computer Art for the Video Picture Wall", 1971 Congress of the International Federation of Inform. Proc.Soc., Ljubljana, Yugoslavia, August, 1971.

4. Dostert, B.H., and Thompson, F.B., "A Rapidly Extensible Language System: REL English". 1969 International Conference on Computational Linguistics, Stockholm, 1969.

5. Dostert, B.H. and Thompson, F.B., "How Features Resolve Syntactic Ambiguity," Proceedings, National Symposium on Information Storage and Retrieval, University of Maryland, April 1971.

6. Dostert, B.H., and Thompson "Syntactic Analysis in REL English", Proceedings, 1971 International Meeting on Computational Linguistics, Debrecen, Hungary, September 1971.

7. "REL Protocol: July 1970" California Institute of Technology, 1970.

Appendix A

REL, July 1970

The following is a copy of an actual printout using the Summer 1970 REL System and illustrating the capabilities of REL English.

Arithmetic

```
What is 234*567-89?
132589
What is 3.45/.003+.56*23?
1162.88
What is 23/2?
11.5
What is 23/3?
7.66666667
What is 23/3000000000?
7.66666667E-08
What is cos(30)*ln(3.45)?
.191021021
_
```

Definitions

```
def:f("8","9"):"8"*"8"+"9"*"9"
What is f(3,4)?
25
What is f(3.5+cos(30),f(5,.04)/2)?
169.6235543
def:x:=variable(9)
def:y:=variable(3+4)
def:g(x,y):f(x,100-y)+x*y
What is g(23,6)?
9503
def:a(x):x+1
What is a(a(a(a(a(a(a(a(a(a(1)))))))))?
11
def:x%y:x*x-y*y
What is 5*(6%7)?
-65
What is g(.45%f(3,4),5)?
396272.929
def:v:23:,:45
What is f(v)?
2554
What is sum f(x,10-x) for x=0(1)10?
770
What is sum product f(x,y) for x=1(2)y for y=3(1)7?
2117717
_
```

Ambiguity

```
def:m:3
def:m:4
What is f(m,6)?
AMBIGUOUS:
(1)      52
(2)      45
What is f(m,m)?
AMBIGUOUS:
(1)      18
(2)      32
(3)      25
def:h(x,y):x+y
def:h(x,y):x-y
What is h(5,9)?
AMBIGUOUS:
(1)      14
(2)      -4
What is h(m,m)?
AMBIGUOUS:
(1)      -1
(2)      0
(3)      1
(4)      6
(5)      8
(6)      7
def:k(x):h(x,100-x)
What is k(5)?
AMBIGUOUS:
(1)      100
(2)      -90
```

Deletion of Definitions

```
What is the definition of h(x,y)?
1.       h(x,y):x+y
2.       h(x,y):x-y
What is the definition of k(x)?
1.       k(x):h(x,100-x)
What is k(5)?
AMBIGUOUS:
(1)      100
(2)      -90
Delete definition 1 of h(x,y)!
What is k(5)?
-90
Delete the definition of h(x,y)!
What is k(5)?
ONE OF THE PHRASES NOT DEFINED ANY MORE.
```

```
What is h(3,4)?
Eh?
What is the definition of h(x,y)?
DEFINITION NOT FOUND.
def:h(x,y):1000-x*y
What is k(5)?
525
What is the definition of h(x,y)?
1.      h(x,y):1000-x*y
—
```

Constants

```
def:c:=constant(34)
What is 2*c?
68
Let c=12!
What is 2*c?
24
What is k(c)?
-56
—
```

Function variables, definition of integration

```
def:f:=function(x):x*x
What is f(5)?
25
def:g:=function(x):3*x+20
What is g(f(5)+k(7))?
1142
def:z:=variable(9)
def:F:=variable(f)
def:inc:=constant(.1)
def:int(F,x,y):inc*(sum F(z) for z=x(inc)y)-(inc/2)*(F(x)+F(y))
What is int(f,0,2)?
2.67
Let inc=.05!
What is int(f,0,2)?
2.6675
What is int(g,5,15)?
500
def:Cos:=function(x):cos(x)
def:Sin:=function(x):int(Cos,0,x)
What is Sin(30)?
-.987825776
What is sin(30)?
-.988031624
—
```

Adding non-numeric data

```
def:Sir Walter Scott:=name
def:author:=relation
def:Waverley:=name
The author of Waverley is Sir Walter Scott!
Who is the author of Waverley?
Sir Walter Scott
def:Edinburgh:=name
def:location:=relation
The location of Sir Walter Scott is Edinburgh!
What is the location of Waverley's author?
Edinburgh
def:arrive:=verb(location,1)
def:depart:=verb(location,2)
def:live:=verb(location,0)
Where did Sir Walter Scott live?
Edinburgh
def:Scott:Sir Walter Scott
def:Billy Scott:=name
def:New York:=name
Billy Scott lived in New York since June 1964!
When did Billy Scott arrive in New York?
 June 1964
def:Scott:Billy Scott
Where is Scott?
AMBIGUOUS:
(1)     Edinburgh
(2)     New York
Where is the author Scott?
Edinburgh
def:boy:=name
Billy Scott is a boy!
Did the boy Scott arrive in New York?
Yes
def:age:=number relation
Age of Billy Scott is 6!
Age of Sir Walter Scott is 56!
Is the age of the author Scott greater than
the age of the boy Scott?
Yes
def:is older than:has age greater than
Sir Walter Scott is older than the boy who lived in New York?
Yes
def:Tom Jones:=name
def:Boston:=name
Tom Jones lived in Boston since May 1954!
Tom Jones departed Boston in October 1967!
Tom Jones is a boy!
Age of Tom Jones is 16!
What was the location of each boy?
Billy Scott      New York
Tom Jones        Boston
What is the average age of boys?
11
_
```

Illustrations of RSL English Capability

Who are the Joneses?
Jack Jones
John Jones
Mary Martin Jones
Who were the Joneses in 1950?
Jack Jones
Jill Jones Smith
John Jones
Mary Martin Jones
When was Jill Jones married?
 from September 1958 to June 1970
When did Jill Jones get married?
 September 1958
When was Jill Jones a Jones?
 from July 1937 to August 1958
Who did Jill Jones marry?
Stan Smith
def:Stan:Stan Smith
def:Jill:Jill Jones
What were the locations of Stan's parents?
New York
Washington
Where were Stan's parents when Stan married Jill?
Washington
Who are Stan's children?
Steve Smith
Sue Smith
Who are the aunts of Steve Smith?
Sally Smith
Sandra Smith Barton
Who are the sons of Steve Smith's aunts?
Bob Barton
When did Sally Smith live in New York?
 from April 1935 to August 1955
 from September 1957 to June 1970
Where was Sally Smith in 1956?
Washington
def:recently:after 1964
Did Jill arrive in New York recently?
No
Did Jill arrive in New York before Stan left New York?
Yes
—

```
Who is Jill's sister-in-law? .
Eh?
def:sister-in-law:sister of spouse
Who is Jill's sister-in-law?
Sally Smith
Sandra Smith Barton
When was Bill Barton's wife Jill's sister-in-law?
 from September 1958 to June 1970
Did Bill Barton live in Washington before Jill's
sister-in-law lived in Washington?
Yes
When was New York inhabited by Jill?
 from September 1955 to March 1961
Was Mary Martin married by John Jones?
Yes
When did Boston cease to be Jill's location?
 August 1955
When did Jill begin to live in New York?
 September 1955
When did Jill have location New York?
 from September 1955 to March 1961
When did Jill have New York as location?
 from September 1955 to March 1961
When did each male Smith live in each city?
Sam Smith        New York          from 1930 to September 1955
                 Washington        from October 1955 to June 1970
Stan Smith       Los Angeles       from June 1961 to June 1970
                 New York          from December 1933 to April 1961
Steve Smith      Los Angeles       from July 1962 to June 1970
When was the number of Smiths equal to 5?
 from August 1938 to September 1956
 from September 1958 to April 1960
Who were Smiths in 1956 and 1959?
Ann Andersen Smith
Jill Jones Smith
 1959
Sally Smith
Sam Smith
Sandra Smith Barton
 from January to September 1956
Stan Smith
Each Smith lived in what city when Jill arrived in Los Angeles?  ***
Each Smith lived in what city when Jill arrived in Los Angeles?
Ann Andersen Smith       Washington
Jill Jones Smith         Los Angeles
Sally Smith              New York
Sam Smith                Washington
Stan Smith               Los Angeles
Sue Smith                Los Angeles
Did all Smiths live in Boston?
No
_
```

```
Did some Smith live in Los Angeles?
Yes
Did some Smith live in all cities?
No
def:eastern city:=name
def:eastern cities:eastern citys
New York, Boston and Washington are eastern cities!
What eastern city was inhabited by each Smith?
Ann Andersen Smith        New York
                          Washington
Jill Jones Smith          New York
Sally Smith               New York
                          Washington
Sam Smith                 New York
                          Washington
Sandra Smith Barton       New York
                          Washington
Stan Smith                New York
Sue Smith                 New York
What Smith lived in some city that was not
an eastern city?
Jill Jones Smith
Stan Smith
Steve Smith
Sue Smith
def:western city:city that is not an eastern city
How many Smiths lived in each city?
Boston            0
Los Angeles       4
New York          7
Washington        4
How many Smiths lived in some western city in March 1962?
3
Did Stan marry Jill, and live in Los Angeles?
Yes
Did Jill liye in New York or Boston?
Yes
Did Sam Smith live in some eastern city and Los Angeles?
eastern city      Yes
Los Angeles       No
What did Stan earn between January 1960 and 1970?
6000 from February to December 1960
8000 from 1961 to 1962
9000 in 1963
10000 from 1964 to 1966
11000 from 1967 to 1969
What were the incomes of the parents of Steve Smith
since Steve Smith was born?
Jill Jones Smith          4000 from July to December 1962
Stan Smith                8000 from July to December 1962
                          9000 in 1963
                          10000 from 1964 to 1966
                          11000 from 1967 to 1969
When did Stan have an income greater than 8000?
 from 1963 to 1969
Did Stan earn more than 8000?
Yes
```

```
Did Stan earn more than 8000?
Yes
Did Stan earn more than Jack Jones?
Yes
When did each Smith earn the maximum income of Smiths?
Sally Smith        from 1967 to 1969
Sam Smith          from 1927 to 1966
What were the incomes of Joneses in 1956?
Jack Jones                 6000
Jill Jones Smith           3000
John Jones                 10000
Mary Martin Jones          0
What was the income of Joneses in 1956?
19000
What was the average income of Joneses in 1956?
4750
What was the maximum income of Joneses in 1956?
10000
What was the sum of the incomes of Joneses in 1956?
19000
Was the income of Jack Jones less than the income
of Stan in 1965?
Yes
Who were Steve Smith's aunts?
Sally Smith
Sandra Smith Barton
What is the definition of aunt?
1.      aunt:sister of parent
What is the definition of sister?
1.      sister:female sibling
What is the definition of sibling?
1.      sibling:child of parent, but not identity
What is the definition of child?
1.      child:converse of parent
What is the definition of parent?
1.      parent:=relation
Who is Steve Smith's grandfather?
Eh?
def:grandfather:father of parent
Who is Steve Smith's grandfather?
John Jones
Sam Smith
Who are Jill's children?
Steve Smith
Sue Smith
def:Ted Smith:=name
Ted Smith is a male!
Ted Smith was born in June 1969!
The parents of Ted Smith are Jill and Stan!
Who are Jill's sons?
Steve Smith
Ted Smith
-
```

Appendix B

Scudder Protocol

The following is a copy of an actual printout using the Summer 1970 REL System made and annotated by Caltech anthropologist, Dr. Thayer Scudder.

The following computer printout has been prepared to illustrate the capacities of a system for relating and simplifying complex data called REL (rapidly extensible language). This experimental system was developed at Caltech by Professor Frederick Thompson and his colleagues with the anthropological data base of Professors Elizabeth Colson (UC Berkeley) and Thayer Scudder (Caltech) used for testing purposes. Their data has been collected over fourteen years (1956-1970) during a systematic long-term study of the Gwembe Tonga in the Republic of Zambia. The vast amount of information collected through time on over 1000 individuals affiliated with six communities can only be adequately studied through computer analysis. To date, however, most software systems have not been designed with the specific needs of anthropologists in mind; in other words, they really have not been user oriented. This is not the case with REL which was designed specifically for behavioral and social scientists.

Though statistical routines have not yet been incorporated in REL, and though only a small proportion of the data has been computerized to date, both of these deficiencies will be corrected during 1971.

The printout is organized to illustrate the use of REL both as a teaching device and as an analytical tool for the principal investigators. Ideally the two investigators would be seated at adjacent consoles so that they could communicate with each other through their data. Two students could also work together in this way. In the present instance they would be told that they were dealing with an African village called Mazulu (after the headman) with a population of 138 in 1956. Using a manual and dictionary on REL English, their assignment would be, for example, to find out as much about the lives of the villagers as they could in a fixed period of time.

One would expect that at first they would ask fairly "safe" questions dealing, for example, with sex ratios, proportion of adults (17 years and older) to children, fertility and so on (pages 1-4). On page 5

the students "discover" that the Gwembe Tonga are polygynous, while clan exogamy and matrilineality come to light on page 9. The main purpose of this printout, however, is to stress the utility of REL for the investigator through a page by page commentary on the printout.

What is the number of total 1957 Mazulu sample?
138
What is the number of total 1957 Mazulu sample who are male?
64
What is (64/(138-64))*100?
86.4864865

def:sex ratio of "Mazulu samples":(number of "Mazulu samples" wh
are male)*100/(number of "Mazulu samples" who are female)
What is the sex ratio of total 1970 Mazulu sample?
Eh?
What is the sex ratio of total 1970 Mazulu samples?
85.6060606

What is the number of adult 1957 Mazulu sample?
56
What is 56/138?
.405797101

What is the number of Mazulu dames57?
31
def:Mazulu crone:=name
All Mazulu dames57 who were born before 1920 are Mazulu crones!
What is the number of Mazulu crones?
9
What is the sex ratio of children of Mazulu crones?
74.2857143

Page 1:

A major attribute of REL encourages the investigator to create definitions as he proceeds. On page 1 sex ratio is defined, while at the bottom of the page a category "Mazulu crones" is defined in such a way as to include only women who have completed their reproductive span. Since any number of such categories can be defined, the investigator can compare, for example, age grades in regard to whatever traits he wishes. Because of the ease with which definitional categories can be created, the user, for the first time, has the opportunity to "play around" with his data -- to look for relationships which have yet to be explicitly conceptualized. This is a tremendous advantage in social anthropology with its incredibly rich informational base. The development of theory in our discipline is difficult, however, not just because of the complexity of human behavior, but also because of the anthropologists' realization that change can be incremental and discontinuous. In the study of societies, the utility, for example, of equilibrium analysis and of comparative statics which has played such a prominent role in classical economics is frankly questionable. At the same time we are beginning to realize that not enough emphasis has been placed on the importance of interaction with close neighbors in influencing marriage choices, cooperation and other institutionalized relationships; relationships which may diverge significantly from stated norms. Against this background, REL facilitates not only the testing of old theories but also, through rapid scanning of much data, the generation and testing of new theory.

Data commentary:

The consistently low sex ratio is puzzling, especially as it relates to the children of Mazulu crones. There are several possibilities, one of which is that elderly women are less apt to report the birth of sons that died as infants. As soon as the data from the other villages is coded, the consistency of this finding can be checked, especially in regard to the sex ratio of the several hundred children in our sample who have been born during the past 14 years.

What is the number of children of each of the Mazulu crones?

Chinyama Mukonka	7
Inkanka Kasamba	8
Kajamba Simakulita	11
Matimba Dobola	10
Mugalabanda Mambo	5
Mwanamoyo Mukamwene	6
Siafunta Mambo	8
Siaguma Siamatwa	9
Silea Chinyama	9

What is the number of children of Mazulu crones, who were miscar
7
What is the number of children of Mazulu crones?
73
What is 73-7?
66
What is 66/9?
7.33333333

What is the number of children of each of the 1970 wives of Mazulu machismos57?

Belita Malala	3
Bwede Muntanga	5
Chifugo Kombo	4
Cigarette Matimba	7
Cigarette Nzala	8
Esnote Mbeza	1
Ester Mioba	8
Fulau Mwindidile	9
Inkanka Kasamba	8
Jessie Bwikili	5
Julia Makuta	2
Kafuele Siakalima	7
Kajamba Simakulita	11
Kasamba Mwiza	4
Malimbu Mulongo	5
Margaret Sialubatwa	5
Maria Chelo	4
Maria Mulongo	3
Maria Mwinde	10
Matimba Dobola	10
Meki Mwemba	8
Melice Siamalenda	8
Moses Munsanje	8
Muchinko Bozi	9
Mudea Sianjungu	9
Munsaka Kalinda	6
Mutibo Mutinta	6
Mwanamoyo Mukamwene	6
Ncengwa Malila	5
Ndeke Manyati	6
Noria Nungu	1
Salah Chibwali	3
Salia Chinyama	2
Salia Makala	5
Siadimba Muleya	3
Siaguma Siamafwa	9
Silea Chinyama	9
Sophia Miyoba	5
Sylea Changu	7
Taulu Musambo	6
Terresa Mukuli	3

When Esnote Mbeza had spouse?
from 1968 to July 1970

When Noria Nungu had spouse?
from September 1968 to July 1970

When Moses Munsanje was born?
January 1934
Who are the children of Moses Munsanje?
Helena Ngoma
Jenny Mwanamoses
Kenl Mwanamoses
Kwale Mwanamoses
Madinga Nelson
Malina Nine
Michel Mwanamoses
Scudder Mwanamoses
Which children of Moses Munsanje have died?
Data not found.

When each child of Moses Munsanje was born?

Helena Ngoma	January 1952
Jenny Mwanamoses	January 1959
Kenl Mwanamoses	January 1965
Kwale Mwanamoses	January 1400
Madinga Nelson	January 1953
Malina Nine	March 1968
Michel Mwanamoses	September 1962
Scudder Mwanamoses	January 1957

When each child of Moses Munsanje was alive?

Helena Ngoma	from 1952 to July 1970
Jenny Mwanamoses	from 1959 to July 1970
Kenl Mwanamoses	from 1965 to July 1970
Kwale Mwanamoses	from 1400 to July 1970
Madinga Nelson	from 1953 to July 1970
Malina Nine	from March 1968 to July 1970
Michel Mwanamoses	from September 1962 to July 1970
Scudder Mwanamoses	from 1957 to July 1970

Pages 2-4:

Demographic analysis of course can be carried out through the use of less flexible and more constrained computer languages. Using Fortran on the batch processor, one of our students has already made a detailed study of the spacing of Mazulu births over the past 14 years. Once a modal interval was determined, attention was focused on the variation. Using least square regression analysis, most of the residual was explained away in terms of three variables. On the other hand, once the proper statistical routines have been added to REL, correlational analysis will be greatly facilitated since it no longer will be necessary to arduously program ahead of time each question.

To anthropologists the ability to have each individual listed by name is a tremendous advantage since the sample is small enough for names to call up a wide range of associations. The problem of privacy, on the other hand, is increased, especially if large numbers of investigators have access to the data base.

On page 4 the year 1400 is a fudge date indicating birth at an unknown time prior to the commencement of our research in 1956. In coding the data we have tried to resist the temptation to estimate birth dates, hence avoiding the very great risk of false concreteness. On the other hand, fudge dates create their own problems so that in the future we will have to estimate dates. The seriousness of this problem fortunately is lessened as our study continues through time simply because revisits at one to two year intervals allow us to give increasingly accurate dates to births, deaths, marriages, divorces and so on. The months January and July are also "fudge months." These will also be eliminated in future coding.

Data commentary:

The consistently high fertility of Mazulu women is apparent, there being no barren women in the sample. Where a complete printout is obtained, as on page 3, it is possible to examine immediately the extremes In this case women with one child had only recently been married.

What is the number of Mazulu machismos57?
25
What was the number of 1957 marriages of Mazulu machismos57?
44
What was the number of 1970 marriages of Mazulu machismos57?
46
def:polygyny rate "in 1957" of "people":(number of marriages "in 1957" of "people")/(number of "people" who were married "in 1957")

What is the polygyny rate in 1970 of Mazulu machismos57?
1.84

What were the villages of major upbringing of 1957 wives of Mazulu machismos57?
chisamu
makololo
masili
mazulu
mukwapata
munjola
nashongo
pembazana
siakaloba
siamwalu
simuchembu
sinafwala
unknown village
Mazulu was the village of major upbringing of which 1957 wives of Mazulu machismos57?
Chitugo Kombo
Impunga Muntanga
Kajamba Simakulita
Kapule Mukuli
Mallmbu Mulongo
Munsaka Kalinda
Siankumba Changu
Sophia Miyoba
What is 8/44?
.181818182

Page 5:

Once polygyny rate is defined, it is, of course, easy to compare polygyny rates of the same aggregate of men at different points in time as well as those of different age categories at any given date.

A logical sequel to the questions dealing with village of wife's major upbringing would be to define the largest indigenous political unit among the Tonga (the cisi or neighborhood) in terms of the specific villages which make it up. For example, def:miyaka cisi:=mazulu pembazana chisamu and siamyalu villages. One could then ask what proportion of 1957 (etc.) wives of Mazulu machismos 57 (etc.) were born in miyaka cisi. In this case the answer would be over 80%.

Data commentary:

Neighborhood endogamy was correlated in 1957 with the control of scarce alluvial gardens by women as well as men. Because of land scarcity, we theorized that the system of land tenure was an incentive for men to marry local women and vice versa. On the other hand, following the relocation of the Gwembe Tonga in 1958 as a result of the Kariba Dam project, the same general pattern of intra-neighborhood marriage continued even though initially the control of land was vested in men because, as pioneers, they had cleared it. In other words, one could also argue that the system of land tenure had virtually nothing to do with intra-neighborhood marriage -- rather that followed from visitation patterns, interaction within the neighborhood being of more importance than relationships to land. The warning here is clear -- interpret statistically significant correlations with care. A great strength of REL is that the investigator can afford, in playing with the data, to search out a multiplicity of relationships whereas in using other techniques he might settle on a single suspected relationship and after lengthy statistical analysis be tempted to read too much into correlations found.

Who are Mazulu machismos57?

Aaron Sipa
Benson Muleya
Emison Mutawajulu
Endeke Siandabile
Felipo Sikalunga
Fremon Masanga
Fulau Sianyandu
Joseph Bebamulilo
Masitut Siakadomwe
Mazambani Sitente
Mazulu Josam
Moses Muleya
Nine Sikalembo
Saugis Siampondo
Senete Sikagoma
Siacaligonya
Sialozi Siabenzu
Siamtunda Mukonka
Siasabe Mazulu
Simon Kayla
Tela Muetwa
Tembo Siankede
Thomas Sikajessie
William Jetulo
Wilson Syamakalu

When each of the Mazulu machismos57 was born?

Aaron Sipa	January 1925
Benson Muleya	January 1937
Emison Mutawajulu	January 1938
Endeke Siandabile	January 1931
Felipo Sikalunga	January 1936
Fremon Masanga	January 1937
Fulau Sianyandu	January 1930
Joseph Bebamulilo	January 1930
Masitut Siakadomwe	January 1930
Mazambani Sitente	January 1929
Mazulu Josam	January 1910
Moses Muleya	January 1939
Nine Sikalembo	January 1923
Saugis Siampondo	January 1919
Senete Sikagoma	January 1905
Siacaligonya	January 1388
Sialozi Siabenzu	January 1905
Siamtunda Mukonka	January 1905
Siasabe Mazulu	January 1939
Simon Kayla	January 1932
Tela Muetwa	January 1932
Tembo Siankede	January 1922
Thomas Sikajessie	January 1912
William Jetulo	January 1928
Wilson Syamakalu	January 1933

Page 6:

Rapid data retrieval is, of course, a major benefit for the computer user. With further programming it will be possible to rank categories of persons not only alphabetically but also by date of birth, date of first marriage and so on.

What are the clans of Mazulu machismos57?
mucindu
muctwa
mukonka
muleya
muyuni
muzamba

The clans of how many Mazulu machismos57 are each clan?

mucindu	2
muetwa	5
mufumu	0
mukonka	4
mukuli	0
muleya	4
mulongo	0
munsaka	0
munsanje	0
muntanga	0
muyuni	8
muzamba	2

The clans of how many Mazulu machismos70s are each clan?

mucindu	4
muetwa	8
mufumu	1
mukonka	5
mukuli	2
muleya	4
mulongo	1
munsaka	1
munsanje	1
muntanga	0
muyuni	13
muzamba	3

What are the clans of Mazulu machismos63s and Mazulu machismos67s?

Mazulu machismos63
mucindu
muetwa
mukonka
mukuli
muleya
muyuni
muzamba

Mazulu machismos67
mucindu
muetwa
mufumu
mukonka
mukuli
muleya
mulongo
munsaka
muyuni
muzamba

What is the number of Sinafwala machismos?
45
What are the clans of Sinafwala machismos?
mucindu
muetwa
mufumu
mukonka
mukuli
muleya
munsaka
munsanje
muyuni
muzamba

The clans of how many 1957 wives of Mazulu machismos57 are each clan?

mucindu	4
muetwa	5
mufumu	0
mukonka	5
mukuli	5
muleya	2
mulongo	5
munsaka	2
munsanje	2
muntanga	4
muyuni	2
muzamba	2

The clans of how many 1970 wives of Mazulu machismos57 are each clan?

mucindu	7
muetwa	7
mufumu	3
mukonka	3
mukuli	5
muleya	3
mulongo	6
munsaka	1
munsanje	2
muntanga	3
muyuni	2
muzamba	1

Pages 7 and 8:

These pages illustrate the importance of REL as a means for generating theory. Among the Gwembe Tonga there are 12 clans. In playing with his data, Scudder noted that six of these were represented in Mazulu village in 1957 whereas 11 were represented in 1970. A quick check showed 7 in 1963 and 10 in 1967 -- a nice progression. Though the population of the village had nearly doubled during this period, by 1968 rapid growth had ended and some fission had begun. Against this background the number of clans within a community may be a possible index for measuring village cohesion. This possibility is reinforced by the situation in Sinafwala village in 1957 (page 8). Ten clans were represented at that time and a period of village fission was just commencing. Once data on clanship and kinship has been coded, REL will facilitate not only correlational analysis but also simulation modeling of the data. On the other hand, the longitudinal nature of the Gwembe Tonga study will help keep the investigator "honest," since probabilistic models can be tested against actual choices made over an extended time period. One intriguing possibility concerns computer modeling of the 1956-57 data in such a way as to replicate behavioral patterns observed in 1963, 1967, 1970 and so on. The next stage would be to project such models into the future bearing in mind, of course, the constant necessity of revising them in the light of new variables and differential interrelationships between old ones.

Data commentary:

Since the founding headman of Mazulu village was a muntanga, the absence of any adult male member from this clan since his death many years ago is of considerable interest. It suggests sorcery. When the investigator has the opportunity to scan large amounts of data, presented in different ways, new notions constantly suggest themselves. At the same time, data inadequacies are made all too clear. This in turn facilitates better information gathering during the next study phase.